



République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Ibn Khaldoun - Tiaret

Faculté des Arts et des Langues/Département de langue et littérature arabe

Laboratoire d'études grammaticales et linguistiques entre patrimoine et modernité en Algérie



Projet de recherche PRFU intitulé :

*Systemes des niveaux de langue arabe depuis l'étude humaine
jusqu'au traitement automatisé*

*Et l'Union des Projets PRFU pour les Etudes Linguistiques affiliés
au laboratoire*

En coopération avec la Faculté d'Informatique et de Mathématique

Organise un Colloque National :

**Épistémologie des structures de la langue arabe entre
les perceptions de l'esprit humain et les acquis des
Grands Modèles de Langage (LLMs)**

Lundi et mardi 18/19 novembre 2024

Amphithéâtre Al-Wiam, de la Faculté de Lettres et de Langues

Argumentaire

À la lumière du développement rapide des technologies de l'information et de la communication, la langue arabe a commencé à retrouver la place qu'elle occupait depuis des siècles, à travers l'adaptation de ses structures phonologiques, morphologiques, syntaxiques, sémantiques, voire lexicales, aux applications et programmes produits par le traitement automatisé du langage naturel *Natural language processing - NLP* -(NLP). Dans plusieurs domaines tels que les chatbots, l'analyse des sentiments, la traduction automatique, la génération de texte, etc.

La raison en est sa nature inhérente, tout à fait compatible avec la logique. Ainsi la langue arabe, dans sa totalité et à tous ses niveaux de structuration, correspond aux existants qui lui sont externes, et c'est une question sur laquelle Ibn Jinni et, avant lui, Al-Khalil ont fait couler beaucoup d'encre sur ses caractéristiques. Ce qui a permis à la machine et ce par le biais d'étiquettes grammaticales et morphologiques limitées d'atteindre une compréhension très performante quant à leur représentation réaliste de l'être humain lui-même, même si elles sont encore pauvres au niveau des bases de données Big Data. Par conséquent on compte beaucoup sur le fait que la langue arabe soit une parmi les langues basiques; pour la simple raison qu'elle est la langue la plus riche en matière de structures diverses qui décrivent des situations différentes avec minutie et dans les moindres détails.

Le progrès humain dans le domaine de l'intelligence artificielle et des techniques d'analyse du Big Data (*Big Data analysis*) vont ouvrir de nouveaux horizons en vue d'une compréhension plus approfondie des structures de la langue arabe, c'est ce qui permet de les analyser de manière aussi bien innovante qu'efficace. C'est pourquoi ce colloque est venu éclairer une deuxième particularité des composantes de la conscience linguistique de la langue arabe, en complétant sa première particularité abordée dans le colloque national précédent intitulé : *Ontologie des structures de la langue arabe entre les fonds philosophiques de l'étude humaine et les algorithmes de traitement automatisé.*

Anthology of the Structures of the Arabic Tongue between the Philosophical Backgrounds of Human Study and Machine Processing Algorithms

Ce dernier colloque, qui a abordé les détails de l'existence physique des concepts au niveau des structures linguistiques de la langue arabe sur les deux niveaux opposés : la conscience humaine et le traitement automatisé des langues naturelles, sera complété par ce deuxième colloque national intitulé

Epistémologie des Structures de la Langue Arabe Entre les perceptions de l'esprit humain et les acquis des Grands Modèles de Langage.

Epistemology of the structures of the Arabic language

Between the perceptions of the human mind

and the outputs of large language models

En abordant mentalement le deuxième volet des conceptions relatives aux structurations de ces structures au niveau linguistique, celui qui éclaire, comme nous l'avons déjà indiqué précédemment, la science de la connaissance de ces dernières dans les détails de ces structures linguistiques de la langue arabe, et ce par le biais d'une étude contrastive entre l'esprit humain et les résultats ou acquis de ce qui est connu dans les cercles de l'information sous le nom des grands modèles de langage (Large Language Models -LLM-) qui traitent le langage automatiquement selon des algorithmes basés sur une logique formelle. Celle-ci n'atteint peut-être pas - du moins à l'heure actuelle - le niveau de la représentation humaine, même si les algorithmes de traitement du langage naturel (NLP) s'inspirent de la compréhension du fonctionnement du cerveau humain. Cependant cette transition de l'esprit vers la machine a en réalité des limites, malgré les efforts insistants des chercheurs de reproduire automatiquement les mécanismes cognitifs. Par conséquent, les capacités de compréhension et de traitement au niveau de la machine restent différentes de celles du cerveau humain. Par conséquent, les capacités de compréhension et de traitement au niveau de la machine restent différentes de celles du cerveau humain.

Mais cela ne nous empêche pas d'essayer de représenter automatiquement les fonctions linguistiques, c'est pourquoi les algorithmes de programmation linguistique nerveuse (PNL) tentent de développer un modèle de traitement du langage naturel dans le cerveau, par exemple pour comprendre le fonctionnement sémantique, la structuration de la phrase et même les nuances minutieuses de la langue. Malgré tout cela les techniques les plus avancées dans ce domaine restent jusqu'à présent incapables de reproduire pleinement la complexité et la précision de la compréhension humaine du langage de façon totale. Cependant cette dernière reste néanmoins une véritable incitation pour les chercheurs à adopter une approche interdisciplinaire dans la conception de leurs projets de recherche visant principalement à résoudre de nombreux problèmes, et favorisant ainsi une compréhension approfondie et globale des enjeux complexes auxquels nous sommes confrontés. A titre illustratif on peut citer l'émergence de recherches pionnières qui tentent de proposer des modèles d'empathie et de les intégrer dans le dialogue pour donner une touche quasi-humaine à la représentation linguistique de la machine, ouvrant la voie à une communication humaine et à une interaction plus sensibles avec les utilisateurs. Ainsi, on peut citer, par exemple, le progrès constaté quant au diagnostic psycholinguistique grâce à l'utilisation des

bibliothèques arabes de PNL telles que 'AraBERT¹ et mBert², qui représentent une étape décisive dans la compréhension des différences linguistiques et des aspects psychologiques de la langue arabe.

Le fait que la technologie n'atteigne pas à l'heure actuelle le niveau de la représentation du langage humain n'empêche pas de continuer à travailler dans ce même domaine. Au moins, il reste du travail pour transférer ce que la technologie a réalisé aux machines pour améliorer les performances neurolinguistiques et des systèmes de programmation. Nous devons réaliser que les capacités des algorithmes sont actuellement dans des limites par rapport aux capacités du cerveau humain si nous voulons construire une machine-homme. Cependant, construire cette machine peut être complètement s'opposer à l'homme quant à sa représentation du langage et de la connaissance. ; Autrement dit, notre perception ne se limite pas nécessairement au processus de clonage. Les traits de cet exploit ne peuvent être réalisés sans les efforts conjugués d'un groupe de chercheurs issus de différentes spécialités, parce que le sujet du colloque est plus œcuménique pour être réduit à la seule spécialité linguistique. Il s'agit plutôt d'un sujet qui occupe une place au sein de ce que l'on appelle les sciences interdisciplinaires. Il a donc besoin des compétences du linguiste, du psychologue, du sociologue, du mathématicien, du religieux, de ceux de l'informaticien. Chacun de ces spécialistes lui apporte ce qui est en relation directe le langage humain.

La phrase prononcée par un enfant de quatre ans a été façonnée par plusieurs milieux ; L'entraînement verbal de sa communication, le suivi psychologique de ses ressentiments et sensations, le lien social quant à son équilibre, la conscience de la valeur de ses biens, et l'alerte morale de son comportement dans un tout fini. C'est pour cette raison qu'il fallait qu'on face appel à tous ces paramètres en vue du traitement automatique de la langue.

De part cette donnée, on peut comparer l'apprentissage de la langue par l'enfant au

¹**AraBERT** developed by Antoun et al.,(2020) is a widely adopted model pre-trained on an extensive corpus of Modern Standard Arabic (MSA) texts. AraBERT is applied in various natural language processing (NLP) tasks, including text classification, named entity recognition (NER), and sentiment analysis (SA) in the Arabic language.

Reference: Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. **AraBERT**: Transformer-based model for Arabic language understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 9–15, Marseille, France. European Language Resource Association.

²**mBERT** released by Devlin et al.,(2019) is a single-language model that was pre-trained using monolingual corpora in 104 languages, including Arabic. This enabled BERT to learn and generalize across multiple languages.

Reference: Jacob Devlin, Kenton Lee, Kristina Toutanova, Jacob Devlin, and Ming-Wei Chang. 2019. **Bert**: Pretraining of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

processus d'entraînement et celui d'utilisation des algorithmes de traitement du langage naturel lui-même, et ce malgré les différences substantielles dans les mécanismes de base du processus d'apprentissage des langues, mais ces deux processus restent deux dénominateurs communs dans la quête de la compréhension et de la production du langage naturel. Cette similitude entre l'homme et la machine pourrait paraître étrange à première vue, mais lorsque l'on suit le processus d'apprentissage du parler par enfant, nous constatons que son cerveau traverse également un processus complexe d'acquisition du langage qui est souvent difficile à comprendre, même pour les scientifiques. Cependant cela ne nous empêche pas de simuler ces processus mentaux au niveau de l'algorithme de programmation neuro-linguistique, à travers plusieurs étapes : exposition à la langue, apprentissage des structures linguistiques, correction des erreurs, développement de la compréhension du contexte et élargissement du vocabulaire et des compétences linguistiques.

Objectifs du Colloque

Ce colloque vise à atteindre un certain nombre d'objectifs dont les plus pertinents sont :

- 1- Analyse des structures de la langue arabe : Comprendre les profondeurs des mécanismes des structures de la langue arabe en tirant profit des théories linguistiques anciennes et modernes par le biais de l'étude des différents genres textuels.
- 2- Comprendre les conceptions de l'esprit humain : étudier la relation entre les conceptions de l'esprit humain et les structures de la langue arabe, et approfondir la compréhension de la manière dont les humains comprennent la langue et ses structures.
- 3- Utilisation des bases de mégadonnées: analyser d'énormes quantités de données linguistiques disponibles en ligne et dans des bases de données linguistiques, dans le but d'en extraire des modèles et de concevoir de nouvelles perceptions sur les structures de la langue arabe.
- 4- Procédures pratiques sur les techniques d'intelligence artificielle : activation des techniques d'intelligence artificielle aux niveaux structurels de la langue, telles que l'apprentissage automatique et l'analyse de texte, pour développer des modèles et des outils permettant de comprendre et d'analyser les structures de la langue arabe.
- 5- Travailler sur la programmation d'algorithmes susceptibles de comprendre les contextes linguistiques : créer des applications intelligentes et des outils efficaces qui aident à une compréhension avancée qui simule l'intelligence par l'homme de la langue humaine, notamment la langue arabe, en vue d'atteindre un certain nombre d'objectifs, tels que l'enseignement et l'apprentissage de la langue, ainsi que la traduction automatique des significations et non celles des mots séparés.
- 6- Accélérer la recherche scientifique dans les institutions académiques pour suivre de près le rythme des développements technologiques hors des murs

de ces institutions: générer des résultats et des conclusions qui contribuent à la recherche scientifique dans les domaines de la linguistique informatique, de la science des données et des applications des derniers à la langue arabe.

Résultats escomptés du Colloque:

1- Parvenir à une compréhension plus approfondie des structures de la langue arabe nous permet d'être capables de conduire celle-ci vers son traitement automatisé de la même manière et la quantité par laquelle les autres langues naturelles ont parvenu à faire de grands progrès et ce par le biais d'une analyse et une description minutieuses et approfondies des structures de la langue arabe au moyen de l'utilisation des plus récentes techniques d'intelligence artificielle et des règles de base de données Big data. Et il n'est pas improbable que l'étude révélera des modèles et des règles linguistiques qui n'étaient pas claires auparavant, et bien particulièrement celle ayant fait l'objet de controverses considérables dans l'ancien cours de linguistique arabe.

2- Nous estimons également de ce travail le développement d'outils analytiques et pédagogiques pouvant être utilisés pour aider les chercheurs et ceux qui s'intéressent à la langue arabe à mieux comprendre les structures de cette dernière.

3- En outre, parmi les résultats escomptés au niveau de la traduction automatique, figure l'utilisation de la nouvelle connaissance au sujet des structures de la langue arabe pour améliorer les performances des systèmes de traduction automatique afin de les rendre plus précis, efficaces et performants.

4- Quant au plan du traitement automatique, l'étude en question estime l'amélioration et le développement des applications relatives à la communication homme-machine, qu'elles soient liées aux domaines de la robotique sociale ou aux applications des constructions linguistiques.

D'une manière générale, l'un des résultats les plus importants attendus de cette étude consiste dans le fait que ce travail ouvrira de nouveaux horizons permettant de comprendre et appliquer les structures de la langue arabe de manière plus approfondie et plus efficace, ce qui contribuera au développement de solutions innovantes dans de multiples domaines.

Axes du Colloques

1. Analyse des structures de la langue arabe :

Étudier les structures linguistiques de base de la langue arabe en analysant les structures phrastiques de l'arabe soutenu, pour identifier celles qui sont fixes et stéréotypées d'une part, et en explorant les modèles et les exceptions d'autre part.

2. Conceptions de l'esprit humain :

Etude des processus linguistiques à l'intérieur de l'esprit humain et celle de la manière dont la structure linguistique est assimilée par rapport à la référence physique externe, en analysant la relation entre le langage et les processus de pensée, de mémoire et d'apprentissage.

3. Analyse des mégadonnées ou des données Big data :

Le travail sur l'organisation des mégadonnées données linguistiques disponibles à travers les thésaurus linguistiques ainsi l'utilisation des techniques de son analyse, telles que l'apprentissage automatique et l'exploration des données pour parvenir à l'architecture des modèles de structuration linguistiques arabes.

4. Applications de l'intelligence artificielle en langue arabe :

Conception de modèles pour l'apprentissage automatique en vue d'analyser les structures linguistiques et appliquer les techniques d'intelligence artificielle et les mettre au service du développement d'outils de traduction automatique, de l'apprentissage en ligne et de l'analyse des textes.

5- Applications pratiques et applications industrielles

Développement d'applications pratiques utilisant les résultats de la recherche pour améliorer la traduction automatique, développer des outils pour l'enseignement de la langue arabe et d'autres applications pour faciliter la communication linguistique entre les humains et les machines.

6- Influence et application sociales

Evaluation de l'impact social et culturel des applications de l'intelligence artificielle dans le domaine de la langue arabe, et ce à travers l'exploration de la manière par laquelle la technologie peut être utilisée pour renforcer la communication et l'intercompréhension entre différentes cultures par le biais de la langue arabe.

Encadrement du Colloque

Président d'honneur du Colloque : Pr. **Dr. BELGOUMANE Berrezoug**, Recteur de l'Université de Tiaret.

Superviseur général du Colloque : Pr. **Dr. ZERROUKI Abdelkader**, Doyen de la Faculté de Lettres et de Langues.

Président du Colloque : Prof. **Dr. BENDJELLOUL Mokhtar**, Directeur du Laboratoire d'études grammaticales et linguistiques entre patrimoine et modernité en Algérie.

Président du Comité du Traitement automatisé : **Dr. OUARED Abdelkader**, MCA. Faculté d'informatique et de mathématiques

Président du Comité Scientifique: Pr. **Dr. BOUHENNOUCHE Fatima**- Membre du Laboratoire des études grammaticales et linguistiques.

Président du Comité d'Organisation et de Suivi linguistique Prof. Dr. BELKACEM Aissa - Chef d'équipe des études grammaticales et morphologiques / Laboratoire des études grammaticales et linguistiques.

Membres du comité scientifique du colloque

Pr. Orabi Ahmed	Université de Tiaret
Pr. Blehocine Mohamed.....	Université de Tiaret
Pr. Benchrif Mohamed	Université de Tiaret
Pr. Bolkhras Mohamed	Université de Tiaret
Pr. Hadouara Omar.....	Université de Tiaret
Pr. HamidaniAissa.....	Université de Tiaret
Pr. Baloul Ahmed.....	Université de Tiaret
Pr. Belkacem Benaouda	Université de Tiaret
Dr. Belkanichi Ali	Université de Tiaret
Pr. Farez Fatima.....	Université de Tiaret
Pr. Mis Souad.....	Université de Tiaret
Dr. Djebali Fatiha	Université de Tiaret
Dr. Haji Zoulikha	Université de Tiaret
Dr. Hadji Amer.....	Université de Tiaret
Pr. Kacem Kada.....	Université de Tiaret
Pr. Karrache Ben Khaoula	Université de Tiaret
Pr. Bouziane Ahmed.....	Université de Tiaret
Pr. Benfriha Djilali.....	Université de Tissemsilt
Pr. Gharbi Bekey.....	Université de Tissemsilt
Dr. Boughari Fatima	Université de Tissemsilt
Dr. Belmihoub Hind	Université de Tissemsilt
Pr. HadouaraMohame.....	Centre Universitaire D'Aflou
Dr. Hamza Boudjmel.....	Centre Universitaire D'Aflou
Dr. Djebari Mohamed.....	Université Khamis Miliana
Pr. Boutouchent El-Ousfoura.....	Université de Blida 02
Pr. Oueldnebia Youcef	Université de Mascara
Pr. Benamar Mehieddine.....	Université de Batna
Pr. Brahmi Ahmed	Université de Djelfa
Pr. Belkhairi Abdelmalek.....	Université de Djelfa
Dr. Bouzidi Mohamed	Université de Tiaret
Dr. Harath Mohammed	Université de Chlef

Comité d'Organisation et de suivi linguistique

Dr. Medjeded Adda.....	Université de Tiaret (Langue Arabe)
Dr. AyedEmel.....	Université de Tiaret (Langue Arabe)
Dr. Qalbaza Youcef	Université de Tiaret (Langue Arabe)
Dr. Djilali Ali.....	Université de Tiaret (Langue Arabe)
Dr. BenssasiBelguendouz.....	Université de Tiaret (Langue Arabe)
Dr. Mesabih Larbi.....	Université de Tiaret (Langue Arabe)
Dr. Ammar Hamou.....	Université de Tiaret (Langue Arabe)
Dr. Oudia Noura.....	Université de Tiaret (Langue Arabe)

Comités Scientifiques Subsidiaires –Chargées du Colloque–

D'après la Spécialité.

Comité de Langue Arabe		
Pr. KerracheBenkhaoula- Tiaret - Pr. Bouzian Ahmed - Tiaret - Pr. Kacem Kada - Tiaret - Dr. Farez Fatima - Tiaret - Pr. Djebali Fatiha. Tiaret - Dr. Hadji Zoulikha. Tiaret - Dr..Boughari Fatima - Tissemsilt - Dr. Belmihoub Hind - Tissemsilt -		
<u>Comité de Français</u> Pr. Malki Benaid- Tiaret Dr. Mehdi Amir -Tiaret Mr. Zouatnia Samir - Khmis- Miliana – Dr. Aounallah Soumia - Tiaret Mr. Benkrama Chahir, Khmis- Miliana	<u>Comité d'Informatique</u> Pr. Chadli Abdellatif - Tiaret Dr. Talebi Omar - Tiaret - Pr. Belarbi Mustapha - Tiaret - Dr. Merati Mdjeddad - Tiaret -	<u>Comité d'Anglais</u> Pr. Benabed Ammar - Tiaret Prof. Hemaïdia Mohamed - Tiaret Dr. Belarbi Khaled - Tiaret - Dr. HemaïdiaGhlamallah - Tiaret -
<u>Comité d'Espagnol</u> Dr.Boubakeur Abed- Tiaret Ms.Oudia Noura, - Tiaret		<u>Comité d'Allemand</u> Dr.Soudani Mohamed- Tiaret Dr. Brahim Khaled - Tiaret

Conditions de participation

- La problématique de l'article doit s'inscrire dans l'un des axes de recherche édictés par le Colloque.
- L'intervention doit être soumise aux exigences de la recherche scientifique académique.
- L'exposé doit être original et non extrait de travaux antérieurs.
- L'exposé doit être rédigée en (Traditional Arabic) 14. Les notes de bas de page et les références dans la même police.
- Les références se situent en bas de chaque page (l'insertion est automatique)
-

Instructions générales:

- Les exposés sont expertisés en double aveugle par un comité scientifique.
- Le Comité Scientifique a le droit de s'excuser de ne pas accepter certaines recherches, sans justifier pour autant les raisons de son refus.
- Le Colloque a le droit de publier les contributions dans une Revue qui lui est affiliée ou dans un livre spécial.
- Le chercheur n'a pas le droit de participer avec la même contribution au Colloque dans une autre manifestation scientifique, ni de la modifier et la publier en tant qu'article.

- Le chercheur est informé de l'acceptation de son article via son e-mail ou son numéro de téléphone.
- Les personnes souhaitant participer doivent remplir le formulaire de participation et l'envoyer, accompagné du résumé et d'un bref CV, au Président du Comité Scientifique via l'e-mail du forum mentionné ci-dessous au plus tard le 28 mai 2024.

Rendez-vous du Colloque

- La date limite de soumission des résumés : **28 mai 2024.**
- La date de réponse aux résumés : **4 juin 2024.**
- La date limite de soumission des recherches complètes : **4 octobre 2024.**
- La date de réponse aux recherches acceptées et d'envoi des invitations : **18 octobre 2024.**
- Rendez-vous du Colloque : **18/19 novembre 2024.**

e-mail : epistemology.structures.ar.18.11.24@gmail.com

<https://chat.whatsapp.com/EMQdSaHt6SR4H5BP5SwPwx>

Formulaire de Participation

La
photographie :
Facultative

- Nom et Prénom(s).....
- Qualifiant scientifique.....
- Grade.....
- Spécialité.....
- Mail professionnel.....
- Téléphone :
- Axe de l'intervention :
- Titre de l'intervention :
- Titre de l'article en anglais :
-
-
-
-
- Résumé de l'intervention.
- Mots clés.....
- Titre de l'intervention en anglais.....
- Résumé de l'intervention en anglais.....
-
-
-
-
-
-
-
-
-
-
-

الذكاء الاصطناعي

تعلم الآلة

التعلم
العميق